Draft. Please ask authors for the latest version

LONG-RUN ASSET RETURNS

David Chambers[a], Elroy Dimson[b], Antti Ilmanen[c], and Paul Rintamäki[d]

7 November 2023

Abstract: The literature on long-run asset returns has continued to grow steadily, particularly since the start of the new millennium. We survey this expanding body of evidence on historical return premia across the major asset classes – stocks, bonds, and real assets – over the very long-run. In addition, we discuss the benefits and pitfalls of these long-run datasets and make suggestions on best practice in compiling and using such data. We report the magnitude of these risk premia over the current and previous two centuries, and we compare estimates from alternative data compilers. We conclude by proposing some promising directions for future research.

Keywords: Asset pricing, historical returns, stock market index, investment management, risk premium, stocks, bonds, real estate, commodities

JEL codes: G11 (investment decisions), G12 (asset pricing), G32 (valuation), G38 (regulation), N20 (financial markets)

Prepared for: *Annual Review of Financial Economics* Volume 16, November 2024 (financial.annualreviews.org).

[a] Cambridge Judge Business School, Cambridge, UK, and Centre for Economic Policy Research, London, UK; email d.chambers@jbs.cam.ac.uk.

[b] Cambridge Judge Business School, Cambridge, UK, and European Corporate Governance Institute, Brussels, Belgium; email e.dimson@jbs.cam.ac.uk.

[c] AQR Capital Management, Greenwich, CT USA; email antti.ilmanen@gmail.com.

[d] Aalto University School of Business, and University of Helsinki, Helsinki, Finland; email paul.rintamaki@aalto.fi.

LONG-RUN ASSET RETURNS

1. INTRODUCTION

This paper surveys the growing body of evidence on the magnitude and consistency of historical returns across the major asset classes of stocks, bonds, and real assets.

There are two main benefits from taking a long-horizon perspective on asset returns (Chambers and Dimson, 2016). The first is simply that it gives us more data at a stroke. Using as much historical data as possible helps investors form better return expectations. Statistically, a longer sample period gives more reliable average return estimates thanks to the law of large numbers and the resultant smaller standard errors. Second, we can test the robustness of in-sample results with the out-of-sample evidence that distant historical data provides. This is especially useful given that accumulating out-of-sample evidence as we move forward is painfully slow in a situation where the in-sample period has been thoroughly data-mined.

At the same time, there are two important caveats to bear in mind when using long-run data. First, the longer the run of financial data, the more likely it is that regime shifts occur which complicate the interpretation of long-run average return estimates. Second, the quality of financial data deteriorates the further we retreat into the past. For example, when stocks and bonds are delisted, it can be impossible for scholars to discover whether the firm was acquired for value or whether it went bankrupt, and investors lost all their money.

Pastor and Stambaugh (2001), Lettau, Ludvigson and Wachter (2008) and others have sought to identify and predict regime shifts. However, paying adequate attention to the quality of long-run historical data is often difficult. In this review, we highlight four common challenges in seeking to uphold data quality, namely, easy-data bias, macro-consistency, replicability, and total-return estimation. At the same time, we make suggestions on best practice when compiling, interpreting, and using such data.

The growth in the empirical literature on long-run asset returns has provided a considerable number of long-run datasets at our disposal. These datasets enable us to address questions using broader and deeper evidence than was possible until recently. The rest of this study deals with five important research questions.

First, we address one of the central questions in empirical asset pricing, namely, whether stocks consistently beat bonds in the long run. Much of this evidence is based on studies of data series starting in 1900, or notably, 1926 in the case of US stocks and bonds - the start date of the CRSP dataset. The choice of start date matters. The evidence regarding average returns drawn from post-1900/1926 studies delivered the conventional wisdom that the equity risk premium over bonds was substantially positive in the US, and pretty much in every other country. However, in returning to this question, we examine the pre-1900/1926 history of stock and bond returns – primarily in the US and UK. The evidence suggests that the gap between the average returns on stocks and on bonds was low in Britain, and seemingly negligible in the US.

Second, we examine just how abnormally low were bond yields in the two opening decades of the 21st century. When viewing the yields of government bonds (or other bonds with low credit risk) in a much longer-run perspective, we conclude that recent yield experience was indeed unprecedented.

The third question concerns corporate bonds and the credit risk premium. The existing evidence on the extent of the excess returns or premia attributable to credit risk is based on a relatively short history stretching back to the 1970s. We therefore ask whether (or not) long corporate bond data histories can provide insights on credit excess returns. In brief, the evidence is at best mixed due to estimation problems and a paucity of data.

Our two remaining questions deal with real assets – real estate and commodities – and how compelling their long-run returns have been compared to equities. When reviewing the evidence on the investment performance of real estate, we discuss the challenges presented by the heterogeneity and immovability of this asset class. The evidence has largely focused on housing and suggests that total returns are below those from equities. We then turn to commodities and critically examine the claims of the recent studies pointing to surprisingly high long-run returns. Here, we conclude that the long-run historical returns on a diversified portfolio of futures have come close to approximating those on equities.

This review begins with a discussion of the challenges in handling long-run financial data. Section 3 then deals with each of our important research questions that can be addressed with historical data. In Section 4, we conclude with some potentially promising directions for future research.

2. CHALLENGES OF LONG-RUN FINANCIAL DATA

Whilst we are proponents of the use of long-run datasets, it is important to exercise care when using such data. The danger of not understanding the limitations of long-horizon data increases the longer the history. There may for example be structural changes in the characteristics of the data such as the number of stocks in the universe and the degree of sector diversification. It is therefore important for historical datasets to include as full a description as possible of the key characteristics of the underlying records and how they were processed.

Below we briefly outline four common challenges when handling long-run historical data. They are easy-data bias, macro-consistency, replicability, and total-return estimation. We focus on stock indices but similar issues arise in other asset classes.

2.1 Biases: Easy-data, survival and success

Researchers can be misled into assuming that long-term historical financial series are of a similar high quality to contemporary indexes. Dimson, Marsh, and Staunton (2002) noted that severe data weakness can result from reliance on easy-to-collect data. Seemingly innocuous decisions, such as selecting a base date that is subsequent to a period of market trauma, can distort estimates of expected return. They highlight the fact that, in 2002, the standard publications on long-term stock market returns in the United Kingdom started at the end of 1918, in The Netherlands in 1947, in Germany in 1952, and in Japan in 1971. They also point towards index compilers' inclusion of markets that had grown, and omission of those that had been laggards. Looking at sixteen standard sources, one for each country, the equity returns over the varying periods covered in prior studies exceeded the returns over the full period starting in 1900 by an average of three percentage points per year. Easy-data bias alone gave rise to estimated long-run equity premia that were almost double the unbiased estimate.

We need to be particularly vigilant about possible biases that could make a dataset unrepresentative of the “true” long-term market experience we are seeking to document. That is, not just adding noise but adding bias. Easy data may expose the researcher to a variety of selection biases both in terms of the securities included and countries studied. One cause of this is that the financial returns of stocks or countries are correlated with decisions about asset inclusion or exclusion.

Survivorship bias may be illustrated through comparisons of equity market country returns. Dimson, Marsh, and Staunton (2021), and previously Brown, Goetzmann, and Ross (1995) as well as Jorion and Goetzmann (1999), stress that US equities have been among the best-performing stock markets over the long term, outpacing the rest of the world by 2–3% per annum. Not coincidentally, the USA has until recently been the primary focus of most academic research. Meanwhile, markets that did not have full capital market histories during the past century and lagged in performance comparisons – including Russia, China, Argentina, and Zimbabwe – have often been excluded from historical studies of global equity market performance. As a result, researchers may have reported an unrealistically favorable picture of long-run equity returns.

Of course, the fact that the USA and UK are recognized as markets with relatively accessible historical data going back to the 1800s and 1900s also reflects success bias. Easy data bias points in the same direction as success bias, and often means that historical samples begin after some difficult periods when asset valuations may have been low. Some studies omit observations that were outliers or uninvestable: periods that were chosen with hindsight. The German hyperinflation of 1923 is one example. Such issues are hard enough to address with recent data; they may be unresolvable with 19th century data.

2.2 Macro-consistency

Early indexes were rarely weighted by the market value of each constituent. They were typically unweighted – computed on an equally weighted basis – or alternatively, like the Dow Jones Industrial Average in New York, were sometimes weighted by the price of each constituent stock. Many other indexes followed investment approaches which could not be replicated by investors, including The Financial Times 30 share index (UK index launched in 1935) and the Value Line index (US index launched in 1961).

In the UK, the FT index is geometric and by construction it must underperform an arithmetic index with the same constituents and weights. Consistent with this, Marks and Stuart (1971) calculated an arithmetic version of the FT index from 1935 to 1970 and found a level that was 38% higher than the actual geometric index. In 1971, a constituent declared bankruptcy and the level of the FT index fell to zero, which was "corrected" by reconstituting the index retrospectively. Value Line's index was also geometric and replicating its performance was impossible, as shown by Brennan and Schwartz (1985), Eytan and Harpaz (1986) and Thomas (2002).

Differing philosophies have been followed in assembling consolidated research databases. The Dimson, Marsh, and Staunton (DMS) approach has been to adopt broad commercial indexes supplemented by peer-reviewed back-histories for each country. Taylor's approach on behalf of Global Financial Data (GFD) has been to follow a sampling procedure based, where possible, on the largest feasible number of stocks in each market – but capped at a maximum of 100 constituents. These two approaches are illustrated in **Box 1**, which reproduces for these two index compilers their descriptions of the respective US equity series. More details are provided in Dimson, Marsh, and Staunton (2023: Chapter 2) and in Taylor (2023).

Box 1: Example of differing philosophies: Best-in-class indexes vs In-house research

DMS: *"The broadest index of early US stock market returns is the one presented in Wilson and Jones (2002), which we use for this study. Our dataset commences with the Wilson-Jones index data over 1900–25. For 1926–61, we use the University of Chicago's Center for Research in Security Prices (CRSP) capitalization-weighted index of all New York Stock Exchange stocks. From 1962 onward, we use the CRSP capitalization-weighted index which includes stocks quoted on all major exchanges including the NYSE, the American Stock Exchange (now NYSE AMEX Equities), and since 1972 Nasdaq."*

GFD: *"We use the GFD-100 Index to calculate returns to stock markets in the United States… The GFD-100 index is revised in January of each year. It uses the largest stocks in the United States from 1792 until 1825, the 50 largest stocks from 1825 until 1850 and the 100 largest stocks from 1850 to date. The index is capitalization-weighted and only includes liquid stocks that trade on a regular basis. Stocks from all regional exchanges and the over-the-counter market are included."*

Indexes can be constructed in many ways. Most useful are indexes that measure the investment performance of a portfolio that contains every security in the investment universe, with each one held in proportion to its outstanding value. Such market capitalization weighted indexes are “macro-consistent” (Sharpe (2010)). They reflect the fact that the average investor must hold the market portfolio (Cochrane (2011)).

When we use a historical index, we therefore need to know how the index is defined. Does it represent a broad all-stock universe or a narrower subset? What has been included, which stocks have been omitted, and why? When there is a choice between broader and more focused coverage, we favor using an all-share index, which provides the broadest coverage in every country. Indexes should wherever possible be value-weighted, not equal-weighted, and index composition should reflect the experience of real-world investors. This preference is challenged by the difficulty of creating stock market indexes for early periods.

2.3 Replicability

Stock market indexes should be designed so that, without any foresight about future performance, investors can replicate index composition with real money. We have already discussed index design and the importance of having a value-weighted arithmetic index rather than, say, a geometric index. We now turn to the challenge of stocks that disappear from databases – such securities often have low returns towards the end of their life. What should a financial historian do to estimate stock market performance when stocks are deleted?

The crudest approach is to exclude stocks that, with the benefit of hindsight, no longer exist. This is equivalent to assuming the investor is clairvoyant – that she exits from a holding before a disaster occurs. When delistings occur, researchers may assign over-optimistic estimates of pre-delisting returns. Grossman (2002) shows a wide range of UK equity return estimates for the period 1870–1913, depending on the delisting return assumptions used. Shumway (1997) shows that, because of delisting bias, estimates of the long-run US equity return, and especially the small-cap premium, were significantly overstated. Foreman-Peck and Hannah (2012) and Campbell, Grossman, and Turner (2021) report that a standard equity index used in the UK had serious omissions that made the index “inadequate” (Gregory (2023)).

Even transitory or occasional omissions matter. Pioneers will frequently see their early work as financial archeologists criticized and their foundational research subsequently improved upon by subsequent scholars. In **Box 2**, McQuarrie (2023) compares his new index for US common stocks in the 19th century with Siegel's (2014) earlier computations. He reveals that one missing stock can have a devastating impact on an estimate of long-run returns.

Box 2: McQuarrie (2023) compares his index composition with Siegel's index

"In the new stock record, the greatest difference from Siegel comes in the decades before the Civil War, and this divergence is readily explained: Siegel's sources omitted the largest single stock that traded before the Panic of 1837, the 2nd Bank of the United States. At the peak before the Panic hit, the 2nd BUS accounted for almost 30% of total market capitalization. It failed spectacularly as the Panic proceeded, with shares dropping in price from $120 to $1.50, and never recovered. To duplicate this omission in the contemporary stock market, it would be necessary to drop Microsoft, Apple, Amazon, Alphabet/Google, and Facebook from the S&P 500; and even these five would not account for as high a percentage of market capitalization as did the 2nd BUS at its peak. Omission of the BUS was the single most glaring error found in Siegel's stock market sources."

Omissions can be numerous. Zweig notes that "More than 300 companies in various industries had been launched in the US by 1801. Most ended up fading into oblivion, leaving no record of their hapless investors' losses. Yet the earliest of the data in *Stocks for the Long Run* tracks only seven bank stocks – all of which were selected because they did not go bust" (Graham and Zweig (2024)). The large number of omissions from the index meant that it failed the test of macro-consistency. Moreover, unless an investor was clairvoyant – knowing which firms were destined to disappear – the investment strategy of the index could not have been replicated in real life.

The expanded interest in research on long-term investment returns, supported by extensive digitization of historical data sources, has given rise to multiple approaches to creating and analyzing security price data. There are single-country studies, such as those mentioned above for the US and UK. More recent initiatives include the work of Annaert, Buelens, and Riva (2016) covering three centuries of Belgian Stock Exchange history and Hautcoeur and Riva's (2018) comprehensive database on the French stock markets since 1795.[1]

[1] See www.uantwerpen.be/scob and https://dfih.fr, respectively.

2.4 Total return estimation

The return on a portfolio held over the long haul is driven by the income it generates. The importance of income was recognized in the first careful study of returns on the New York Stock Exchange by Fisher and Lorie (1964, 1968) who computed not only capital appreciation but also total returns including dividends. Ibbotson and Sinquefield (1976) also noted the importance of dividend income in their landmark study of US stock returns. In general, as we go back in time asset price data are more easily available than dividend data, and so price-only return estimates tend to be more readily compiled than total return estimates. Good estimates of dividend income are difficult to assemble, especially for intervals from long ago and for markets with incomplete records.

Jorion and Goetzmann's (1999) analysis of investment returns from stock markets around the world was based on long-term capital appreciation indexes in 39 countries. They note that the inclusion of dividend income would elevate return estimates in both the USA and other countries. Moreover, small adjustments to estimated dividend yields have a sizeable impact on long term returns. Zweig (2009) highlighted inaccuracy in dividend estimates in Siegel's (1992) early data and the resulting upward bias in computed returns.

When we consider pre-1900 data, income return dominated price appreciation in the total real returns of stocks, bonds and real estate. However, income or payout data is harder to source. Collecting data on realized as opposed to contractual income payments becomes particularly challenging in the case of coupons on bonds and rents on properties. In the latter case, and in the absence of observing transacted rental yields over time, Jordà, Knoll, Kuvshinov, Schularick, and Taylor (JKKST) (2019) relied on back-casting from relatively recent yields. In the case of bond defaults and of vacated properties, identifying recovery rates and the length of rental voids may not be possible. Plugging in the contractual cashflows or interpolating data will overstate realized returns.

Finally, we close this section with some remarks on best practice, in the spirit of Annaert, Buelens, and Riva (2016) and Baltussen, van Vliet, and van Vliet (2023). Scholars should do the best they can with the available historical data, with all its inherent problems and shortcomings. At the same time, there should be a clear disclosure of data sources and design and measurement choices, whilst highlighting challenges, potential biases, and interpolation or back-casting of data. Good practice involves robustness checks of reported results reflecting the different possible design choices. Last and not least, we should give full credit to earlier researchers, especially those who have undertaken the painful effort to create and clean data and to make it available to others.

3. THE BIG QUESTIONS

With long and broad datasets, we can take a fresh look at many issues. Our choice of questions discussed below represent those which the authors believe are of first order importance to academics and practitioners alike, and which have been directly addressed in the prior literature.

3.1 Equities: was the 20th century an anomaly?

An abundance of evidence has shown that the total returns of equities have exceeded those of (government) bonds over various long-run periods since the beginning of the twentieth century – both in the US (Ibbotson and Sinquefield (1976), Siegel (1994, 2022)) and in the rest of the world (Jorion and Goetzmann (1999), Dimson, Marsh, and Staunton (2002)). First, we address one of the central questions in empirical asset pricing, namely, whether stocks consistently beat bonds over the long run. In large part, this evidence relies primarily upon data series starting in 1900, or in the case of the United States, 1926 when the University of Chicago's CRSP (The Center for Research in Security Prices) data starts. In this section, we first discuss the history of stock and bond returns before 1900/1926 and we critically review out-of-sample estimates of the equity risk premium before making comparisons with twentieth century estimates. As well as looking at the US, we consider Britain – the home of the premier stock and government bond markets in the nineteenth century.

Below we review the long-run evidence on the total returns achieved by stocks. The ideal measure of stock returns is the compound average total return, including price changes and any payouts to existing shareholders, for a market-capitalization-weighted universe of all listed stocks, adjusted for delistings and corporate actions.

Researchers have undertaken admirable detective work in physical and digital archives to gather data and construct estimates of nineteenth century stock returns. In the US, the first pre-CRSP evidence on stock returns drew on the efforts of Smith and Cole (1935), Macaulay (1938), and Cowles (1938). These estimates were later used by Schwert (1990) and Siegel (1992) among others as their pre-CRSP source, and subsequently modified by Goetzmann, Ibbotson, and Peng (2001). Most recently, Baltussen, Van Vliet, and Van Vliet (2023) have made a detailed and valuable contribution to our understanding of US stock returns from 1866 to 1926 based on a careful data collection exercise. McQuarrie (2023) has made a similar heroic effort going back to 1800 for the US.

Similarly, in the UK, considerable efforts have been made to extend stock return histories back to 1829 (Acheson, Hickson, Turner, and Ye (2009), Campbell, Grossman, and Turner (2021)), and to preceding periods (Campbell, Quinn, Turner, and Ye (2018)). We have used an even earlier equity index (Gayer, Jacobson and Finkelstein (1940)) with dividend information from GFD to extend the Campbell, Grossman, and Turner (2021) indexes back to 1800. Thomas and Dimsdale (2018) provide the corresponding bond returns and inflation rates, reflecting the adjustments proposed by Klovland (1994).

Table 1 summarizes the annualized nominal returns of equities and bonds, the equity risk premium (over bonds) and the annualized real returns of equities over 1800-99, 1900-99 and 2000-22. Panel A covers the US returns reported by Siegel and later reconstructed by McQuarrie (2023), accompanied by estimates from Taylor (2023) and Dimson, Marsh, and Staunton (2023). Panel B provides UK estimates for 1800-99 for equities (primarily from Campbell, Grossman, and Turner (2021)) and bonds (from Thomas and Dimsdale (2018)), accompanied by estimates up to 2022 from Taylor (2023) and Dimson, Marsh, and Staunton (2023). All the quoted figures are geometric means (compound returns) and thus lower than arithmetic means (simple returns), especially for the more volatile series.

The main message is that the US equity premium over (government) bonds at 5% and above in the 1900s was considerably above estimates for the 1800s ranging between +1.6% (Siegel) and –0.6% (McQuarrie). The disagreement between these two figures is due to differing estimates of equity and bond returns in roughly equal measure. First, Siegel overestimated equity returns and (as noted above) his estimate suffers from both survivorship bias and a crucial stock omission. Second, there are problems with Siegel's attempts to proxy risk-free bonds in the 19th century. McQuarrie eschews any attempt at such a proxy in favor of estimating the return on "a hypothetical bond index fund" (McQuarrie (2021)) drawing on a comprehensive dataset of municipal, federal, and corporate bonds. This inevitably will capture some credit premium in municipal and corporate bonds, and not a pure riskless return, partly explaining McQuarrie's higher bond return and lower equity premium in the early data. (All bond sources use long-dated government bonds since the 1920s, though with different maturities.) Hence, the 1800-99 US equity premium over truly riskless bonds is probably not a negative figure but most likely a small positive number and still a lot smaller than the premium for the 1900s.[2]

The UK evidence in Panel B is broadly consistent with this picture. The DMS+ series merges Campbell, Grossman, and Turner (2021) and Thomas and Dimsdale (2018) estimates for 1800–99 with Dimson, Marsh, and Staunton (2023) estimates for 1900–2022. The UK equity-bond premium for the 1800s is estimated at 1.4% (DMS+ and Taylor). We consider the UK estimate for the 19th century to be a more reliable measure of the equity premium than the US estimate, given that the bond returns are based on obligations of the British government (so-called Consols) – the prime credit available to investors throughout the 1800s. The 223-year estimate of the annualized UK equity-bond premium is in the range 2–3%, which is again well below the 20th century premium of over 4%.[3]

[2] Difference-in-means tests show that the equity-bond premium was significantly lower in the 1800s than the 1900s, with t-statistics ranging from -2.1 (Siegel) to -3.6 (McQuarrie).

[3] There are fewer sources of short-term bill returns than long-term bond returns before 1900, so we do not include them in Table 1. GFD estimates a 1.2% lower return for US bills than bonds in the 1800s, reflecting a healthy term premium (compared to a term premium of 0.7% in the 1900s and 2.4% in the 2000s). The term premium was lower in the UK: 0.2% in the 1800s, near 0% in the 1900s, and near 2% in the 2000s.

Table 1. Annualized Percentage Returns on Stocks and Bonds since 1800 (%)

Panel A: USA	Siegel	McQuarrie	Taylor	DMS+	Siegel	McQuarrie	Taylor	DMS+
	US Equities Nominal Return				US Bonds Nominal Return			
1800-99	6.8	5.5	6.8	5.5	5.1	6.2	5.5	6.2
1900-99	10.5	10.2	11.2	10.3	4.8	4.9	4.4	4.5
2000-22	6.7	6.1	5.5	6.2	5.2	5.2	4.0	5.3
All	8.4	7.7	8.6	7.7	5.0	5.5	4.9	5.3
	US Equity-bond Premium				US Equities Real Return			
1800-99	1.6	–0.6	1.2	–0.6	7.2	5.9	6.7	5.9
1900-99	5.5	5.1	6.5	5.5	7.3	7.0	7.8	7.0
2000-22	1.5	0.9	1.5	0.9	4.1	3.6	3.0	3.6
All	3.3	2.1	3.6	2.2	6.9	6.1	6.8	6.2

Panel B: UK	CGT & GK	Taylor	DMS	DMS+	CGT & GK	Taylor	DMS	DMS+
	UK Equities Nominal Return				UK Bonds Nominal Return			
1800-99	5.4	5.5		5.4	4.0	4.0		4.0
1900-99		10.1	10.2	10.2		4.9	5.4	5.4
2000-22		4.5	4.4	4.4		3.7	3.9	3.9
All	5.4	7.4	9.1	7.4	4.0	4.4	5.1	4.6
	UK Equity-bond Premium				UK Equities Real Return			
1800-99	1.4	1.4		1.4	5.4	5.6		5.4
1900-99		5.0	4.6	4.6		6.0	6.1	6.1
2000-22		1.1	0.5	0.5		2.6	1.9	1.9
All	1.4	2.9	3.8	2.7	5.4	5.3	5.3	5.5

Sources: **Panel A:** McQuarrie (2023), Taylor (2023), Dimson, Marsh, and Staunton ("DMS") (2023), and the authors of this paper. McQuarrie (2023) reconstructs Siegel data from original Smith and Cole, Schwert, Goetzmann, Ibbotson and Peng, Cowles, and CRSP (post-1926) data. **Panel B**: Golez and Koudijs ("GK") (2018) over 1800-29, Campbell, Grossman, and Turner ("CGT") over 1830–99. Bond returns and inflation are from Thomas and Dimsdale ("TD") (2018). The DMS+ series merges the CGT & GK data up to 1899 with DMS estimates for 1900–2022. All returns are annualized geometric means in local currency.

Recent empirical studies of equity and bond returns in the 19th century for Belgium and France – both were among the most important capital markets in the world at this time – also indicate a low premium compared to the 20th century. Annualized nominal returns for Belgian equities and government bonds are estimated as 4.0% and 4.4% respectively over 1838–1900 (Van Mencxel, Annaert, and Deloof (2022) Table 5). This slightly negative equity-bond premium compares with 3% for 1900-99 (Dimson, Marsh, and Staunton (2023) Table 24). In France, annualized nominal returns over 1854-1914 for equities were estimated at 5.3% and for government bonds at 4.0% implying an equity-bond premium of 1.3% (Le Bris and Hautcoeur (2010) Table 10). This compares with 4.3% for 1900-99 (Dimson, Marsh, and Staunton (2023) Table 37). In short, the robust and substantially positive equity premium of the 1900s – almost pervasive across countries in the Dimson, Marsh, and Staunton dataset – was much smaller and quite close to zero in the previous century.

Whilst a full discussion of the fundamental factors driving this "underperformance" of the 19th versus the 20th century is beyond the scope of this review, we highlight two important factors here. First, the 1800s experienced both lower growth in per capita income and lower productivity growth than the 1900s. This reflects the comparatively modest payoff of the first industrial revolution (steam) compared to that of the second industrial revolution (electricity and the combustion engine) of the early 20th century and the third (information and communication technology) revolution of the later 20th century. Second, the great improvements in corporate governance and disclosure so necessary to induce outside equity ownership did not occur until the 20th century. These improvements in corporate governance contributed to the general rerating of equities and to stock returns outpacing bonds over the 20th century. Relatedly, bonds were hardly riskless in the 1800s.

A decomposition analysis of total equity returns supports such claims about the relatively poorer fundamentals of the 1800s. The components of total equity returns in real terms are the dividend yield, the real growth in dividend per share, and the change in equity valuation. The decompositions for the US and UK in the 1800s and 1900s show that dividend yield is overwhelmingly the most important contributor to returns in both centuries.[4] The real growth in dividends and the upward trend in equity valuations of the twentieth century were nowhere apparent in the earlier period.

To sum up, the equity risk premium was exceptionally high in the 20th century, while being low in the 19th century – notably in the USA. Both centuries may have been anomalous. Consistent with this, leading experts at a recent equity premium symposium estimated a prospective equity-bond premium of between 0% and 5% (Siegel and McCaffrey (2023) Appendix A). The best guidance on future premia may be between the experience of the lucky 1900s and that of the disappointing 1800s.

---

[4] Total equity return can be decomposed in increasingly granular ways. The most common decomposition is in the spirit of the dividend discount model, splitting the return into dividend yield (d/p), dividend-per-share growth, and valuation (p/d) change components (e.g., Campbell (2018)). Dimson, Marsh, and Staunton (2023: Table 12) present this decomposition for 23 national markets over the period 1900–2022. Kvshinov and Zimmermann (2022) document the declining role of dividend income in stock returns across centuries.

3.2 Bonds: How abnormal were the recent low bond yields?

Having looked at bond returns in the previous section, here we consider bond yields in a long-run perspective. The finance literature has anchored on the fact that recent bond yields appeared abnormally low compared to the past and that the common investor response was to reach for yield.[5] Certainly, bond yields in the past decade were as low as they had been since 1900 (reaching negative real yields in major bond markets, even negative nominal yields in Japan, Germany, and the UK). Given that bond markets have been around for a lot longer than stock markets, what does the pre-1900 data tell us about just how abnormally low real yields became?

The longest data sources relating to loan or bond yields are Homer and Sylla (2005) and Schmelzing (2020). The former dates back to ancient times and the latter to the early 14th century. However, interpreting these estimates of historical bond yields is problematic. Bond contracts in the nineteenth century incorporated many features which make estimating a consistent series of bond yields very difficult. Beyond maturity and coupon, there are complications such as sinking funds and embedded options (mostly for the issuer's benefit), tax features, and even lotteries. Furthermore, bond ratings for either government or corporate bonds were not introduced until the 1900s making credit risk largely impossible to estimate before then.

Notwithstanding such challenges, Schmelzing (2020) – through his own efforts as well as those of others – assembles a global history of yields on loans or bonds since the early 1300s. We focus on GDP-weighted average yields across eight countries once each became available: Italy 1310, Great Britain 1310, Germany 1326, France 1387, Spain 1400, Netherlands 1400, USA 1786, and Japan 1870. Among the six European states, only France existed when the time series began, so the loans in the early sample were to local rulers, princes, or cities. As capital markets evolved over time, loans were replaced by tradeable bonds, and for the past century, the dataset covers

[5] Historical analysis shows that there was nothing new in the 2010s reach-for-yield activities. This is what investors have done repeatedly when yields fall to abnormally low levels. Korevaar (2023) covers Dutch investors in the 1600-1700s and Turnbull (2017) British investors in the 1800-1900s. Chancellor (2022) quotes Walter Bagehot most memorably from the 1850s: "John Bull can stand many things, but he can't stand two percent."

10-year government bonds. Earlier, loan/bond maturity and credit quality varied over time and across countries.

Table 2 tracks the evolution of nominal and real (inflation-adjusted) yields as well as inflation and real GDP growth. (The last series is from the Maddison-Groningen database, to be discussed below.) The table shows that nominal and real yield were in double digits in the Middle Ages, but fell to about 5% near the Industrial Revolution, and ever lower thereafter.[6] Global inflation was mildly positive but near zero in most centuries before the 1900s. Centennial average inflation near zero did not mean stable inflation near zero, as there were oscillating waves of inflation and deflation.

Table 2. Centuries of Declining Nominal and Real Bond Yields (%)

	1314-99	1400-99	1500-99	1600-99	1700-99	1800-99	1900-99	2000-18
Nominal rate	11.5	12.2	9.2	7.4	5.1	5.4	5.4	3.0
Real rate	10.4	11.5	7.1	6.8	4.2	4.9	1.0	1.3
Inflation (7y, centered)	1.2	0.8	2.1	0.6	0.9	0.5	4.4	1.7
Real growth (GDP %)	0.4	0.1	0.2	0.2	0.2	0.9	2.2	1.1

Sources: Schmelzing (2020) and Maddison-Groningen (2020). Notes: All series are GDP-weighted averages across countries and arithmetic means across years of annualized rates from (up to) 8 countries since data are available, in most cases before the birth of the nation. Nominal rate refers to loans or bonds, real rate is the difference between the nominal rate and the 7-year centered average inflation rate. The last series, real GDP growth per capita, is from the Maddison-Groningen database, and where possible reflects the same GDP-weights as the Schmelzing series.

Looking at Schmelzing's data in Table 2 reveals that before the twentieth century inflation was very low and there was little difference between nominal and real yields which both fell to 3-4% around 1900. The twentieth century saw more persistent price rises, particularly the Great Inflation of the 1970s which pushed nominal rates up to double digits – a level last seen in the 1500s. The following Great Moderation after 1982 witnessed global convergence in inflation rates toward 2% – back to the low levels which characterised the pre-1900 economy. Even after inflation expectations stabilized at low levels in the 2000s, real yields and policy rates kept falling to negative territory. Eventually, even nominal yields in many European countries and Japan turned negative, or reached record low levels in places where they didn't (the US and the UK). The global rise in inflation in 2021-22 finally turned bond yields higher after a 40-year downtrend. The yield evidence in Table 2 stretching back over centuries provides confirmation that the recent low (even negative) bond yields were truly

[6] Looking even further back, Homer and Sylla document yields near 20% in Mesopotamia four or five millennia ago. A local trough occurred amid Pax Romana near CE1 with maybe 4% yields, before double-digit yields again became commonplace.

exceptional.

Alongside real bond yield data, we have very long run estimates of income per capita growth across countries assembled by the Groningen Growth and Development Centre; see Maddison-Groningen (2020). Classic economic theories would suggest that equilibrium real yields and real growth should be similar. However, very long-run empirical evidence suggests the opposite.[7] Real economic growth was 0.1-0.4% per year before the Industrial Revolution, whereas the past two centuries have seen about 1-2% real annual growth in advanced economies (last row of Table 2). Thus, the recent two centuries of relatively fast economic growth and higher inflation coincided with low real yields, while the earlier centuries of slow growth and low inflation witnessed much higher required yields.[8]

This apparent puzzle of high real interest rates in a slow-growth world may be explained by capital scarcity, the impatience of savers at a time when most people lived near subsistence level, as well as high financial intermediation costs (Bernstein (2022)). Most economists studying the fundamental determinants of real interest rates have focused on the expected consumption growth rate rather than time preference. Whilst this is appropriate in the modern world, it seems less so in pre-industrial times.

3.3 Do long histories give us clear evidence on excess returns to credit risk?

We next turn to corporate credits. The credit premium is defined as the (expected or realized) excess return of corporate bonds over comparable government bonds. The same data issues around accurately capturing the various features of bonds discussed above in the context of bond yields make it nearly impossible to measure the credit premium in early corporate bond markets.

The financing needs of the railroads stimulated the growth of large corporate bond markets from the mid-19th century onwards in the US and the UK. Consequently, there

---

[7] From Wicksell (1898) to Woodford (2003), the natural (equilibrium) real rate is tied to the productivity of capital, serving as one key anchor to monetary policy rules. In neo-Keynesian dynamic stochastic general equilibrium models, this rate reflects productivity growth and investor time preference. Consumption-based asset pricing links the riskless real rate to consumption growth (as well as time preference and Jensen's inequality terms); see Campbell (2018).

[8] This evidence contradicts Piketty's (2013) claim that the rate of interest (or broader investment return) always exceeds the rate of economic growth, "R>G". We bypass this topic because we could debate endlessly about which interest rate and which growth rate to use.

are a few studies that add to our knowledge in this area. Giesecke, Longstaff, Schaefer, and Strebulaev (2011) estimate annual value-weighted percentage default rates for US non-financial corporate bonds over 1866-2008. The authors report the annual time series of default rates and observe a decline over time with rates averaging 4%, 1.4% and 0.3% for 1866-1899, 1900-45 and 1946-2008 respectively. To estimate the credit spread, they utilize Homer and Sylla's yield data on "high-grade" railroad bonds and "high-grade" New England municipal bonds for 1866-1914 and "prime" corporate bonds and long-term government bonds after 1914. These category definitions can only crudely take account of credit differences across corporates in the pre-bond-rating era. We also again note the problems of proxying US government bond yields pre-1914. With this data, the authors report an average credit spread of 1.5% for the entire sample period 1866-2008 (without disclosing the annual time series). The average default rate is also 1.5%. Given their "broad-brush" assumption of a 50% average recovery rate of defaulted losses, they estimate the average credit risk premium to be around 0.75%.[9] This figure compares with an average credit premium of 1% for investment grade corporates from 1973 to 2020 (Ilmanen, 2022).

Coyle and Turner (2013) report annualized real returns on UK corporate bonds over the period 1861-2002 of 2.4% and a credit spread of approximately 1% versus government bonds. The spread was higher at 1.7% for 1861-1913 than afterwards. Since ex post default losses were negligible, the credit premium is of a similar magnitude. As with the US, however, there is no available information on corporate credit quality in the 19th and early 20th centuries. Given the sample comprised all corporate bonds traded on the London stock market for which listing requirements were relatively strict, we might infer that credit quality of this bond sample was generally high.

A further data point on corporate bonds is provided in the recent study by Van Menxcel Annaert, and Deloof (2022) on Belgium from 1838 to 1939. As mentioned above, the Brussels stock market was relatively large in the 1800s and there is an abundance of data. This study reports overall a small positive corporate credit spread over

[9] Giesecke, Longstaff, Schaefer, and Strebulaev (2011 p.243) cite research by Hickman (1960) which implies an average recovery rate of 62.5% of par value for 1900-44.

government bonds of around 0.4%. There is no discussion of default losses and recovery rates to enable us to estimate the credit premium.

Overall, we can only measure corporate credit spreads and credit premia quite crudely. In the absence of bond rating data for corporate bonds before the 1920s, it seems difficult for scholars to be able to provide more precise estimates of the corporate credit premium than the few we have. The existing sparse evidence suggests that there was a modest credit premium in the 19th as well as the 20th century. Given that Kizer, Grover, and Hendershot (2019) and McQuarrie (2020) find problems even with the US corporate bond return data in the early decades after 1926, it is likely to remain the case that the most reliable evidence on the credit premium begins only in 1973.

Finally, we do have an estimate of the sovereign credit premium. Meyer, Reinhart, and Trebesch (2022) have estimated a credit spread before default losses of 4.3% on foreign-currency government bonds between 1815 and 2016 over UK and US government bonds. This spread was about 1% lower for the first hundred years, 1814-1914. In credit events (sovereign debt restructurings), the typical realized haircut was less than half, implying an average annual default loss near 1%. The resulting average ex-post sovereign credit premium of around 3% is considerably above that for corporates.

3.4 Real estate: Do housing returns match equities in the long run?

Much of the academic literature regarding real estate performance has focussed on housing given its importance within household wealth. There are numerous historical studies of house price changes over time.[10] However, very few attempt to estimate *total* returns, including rental yield net of costs, to housing over long time periods largely due to the data demands of making such estimates. Total return is the focus of the discussion in this section.

There are two main types of historical data which are employed to analyze long-run housing returns. "Macro" studies seek to exploit the aggregated national statistical

[10] The historical US house price series of Shiller (2005) are the best known but have been subsequently improved upon by for example Lyons et al (2023).

index series of house prices and rents. Jordà, Knoll, Kuvshinov, Schularick, and Taylor (JKKST) (2019) collected much of this macro data in their impressive multi-country study of housing returns (alongside stocks and bonds) from 1870 to 2016.[11]

In contrast, "micro" studies seek to exploit detailed, property-level transaction prices, realised rents and the associated costs to the landlord typically in a specific location. Such archival evidence of the required quality and without discontinuities is harder to source for properties than for stocks or bonds. As a result, there have been only a handful of such micro studies of total returns to date – Paris over 1809–1943 and Amsterdam over 1900–1979 by Eichholtz, Korevaar, Lindenthal, and Tallec (EKLT) (2021), and England over 1901-1983 by Chambers, Spaenjers, and Steiner (2021).

There are five important differences when estimating total returns on housing compared to publicly traded stocks and bonds. First, the price formation and trading process of real estate is different. Trading is infrequent, and market values are much more difficult to establish consistently. Hence, investment returns can only be estimated with noise.

Second, (residential) real estate prices reflect *private value*, namely, a use value derived from ownership which is nonmonetary and nontradable.[12] It is private in the sense that it depends on the identity of the owner. For private-value assets, any two potential buyers will be willing to pay different amounts—reflecting differences in preferences and relative wealth—even when they have identical resale strategies and agree on future monetary cash flows. Because of the illiquidity of the markets in which these assets are traded, variation in private values can translate into systematic differences in transaction prices and thus financial returns between market participants.

Third, *quality improvements*: a new three-bedroom house today is built to a higher standard than one built a hundred years ago in the same location. Hence, we need to adjust for such quality improvements when estimating returns – both price changes and income. It is noteworthy that EKLT (2021), who address the issue of quality adjustments head on by using new historical data, conclude that "most of the increase

[11] The authors provide in www.macrohistory.net a free multi-country dataset of their collected asset class returns and select macroeconomic series since 1870.

[12] Goetzmann, Spaenjers, and van Nieuwerburgh (2021) review the literature on other private value assets including real estate.

in housing expenditure that did occur is attributable to increasing housing quality rather than rising rent."

Fourth, investors must bear the substantial costs of maintaining their properties to realize capital gains in line with a quality-adjusted price index. Yet, prior studies do not always adequately take such costs into account.

Last, we need to be aware of the so-called "superstar city bias" in the case of housing (Gyourko, Mayer, and Sinai, 2013). Prior historical studies—even of "national" housing prices—with a focus on capitals and other large cities are thought to exhibit an upward bias in their average rate of house price appreciation. On the other hand, a very recent and extensive long-run study claims that when we estimate total returns this bias reverses due to lower city rents (Amaral, Dohmen, Kohl, and Schularick, 2023). The higher total return in rural housing is perhaps compensation for the increased risk in the form of higher co-variance with income growth and lower liquidity.

When estimating housing returns therefore, scholars need to be particularly mindful of the last three of these real estate characteristics. Both prices and rents need adequate adjustment for housing-quality improvements, maintenance and other costs need to be fully accounted for. The superstar city debate highlights the importance of assembling income and capital appreciation with particular care.

Quality-adjustment is usually addressed by estimating price appreciation using repeat sales and rental changes of the same set of properties. This obviously requires a very large sample of properties to generate sufficient transacted price observations given the infrequency of trading. It is possible to use smaller samples along with hedonic regressions, provided there is sufficiently detailed data available to control for the heterogeneity in property characteristics.

When estimating income returns, there are two further considerations to be kept in mind. We require data on *realised* (as opposed to *contractual*) rents that take account of any discounts or voids. Furthermore, income yields need to be estimated from prices and rents observed concurrently on the *same* set of individual properties. Where such a time series of realised income yields on a set of properties cannot be estimated, scholars have resorted to starting with a contemporary rent-to-price ratio and then extrapolating backwards using historical price and rent index changes. This methodological approach risks compounding measurement errors over any long-run period. According to EKLT (2021 p.3629), "it is difficult if not impossible to accurately derive the evolution of rental yields over time from series of house prices and rents that relate to different sets of dwellings or use different methodologies".

Turning to the empirical evidence, the JKKST (2019) study asserted that on average housing delivered a total return which at least matched that of equities.[13] Housing performed better than equities in 10 (not including the US or UK) out of the 16 countries in the JKKST sample over 1870-2015. Across all countries the *equal-weighted* real return was 7.3% on housing compared to 6.7% on stocks. When switching to GDP-weighting, housing (6.7%) performs slightly worse than equities (7.0%).

Dimson, Marsh, and Staunton (2018) were among the first to challenge this surprising conclusion in their study of broad house price indexes over the period 1900–2017 in eleven national markets. After adjusting for population size, supercity bias, maintenance and insurance costs, and quality improvements, they estimated annualised house price returns were negative.

Turning to total returns, the micro housing studies of France, Holland and England report lower estimates than JKKST (2019); see EKLT (2021) and Chambers Spaenjers, and Steiner (2021). **Table 3** summarises the estimates of capital gains and of total returns in real terms for all three studies. Whilst JKKST estimate country-wide returns, the micro studies estimate returns for England, Paris, and Amsterdam respectively. The Paris and Amsterdam samples are sufficiently large to yield enough repeated price observations for the estimation of capital gains by using repeat-sales regressions.[14]

Each of the three micro studies report total return estimates substantially lower than those of JKKST. This most likely reflects the difficulties in a macro study arising from not fully adjusting for quality improvements and fully accounting for all costs, from the mis-estimation of realised income yields by extrapolating backwards using different

---

[13] JKKST also reported lower volatilities for housing compared to equities and hence higher Sharpe Ratios. Given that the vast majority of retail investors in housing would be unable to diversify their housing exposure beyond a single property or outside their home market, such low levels of risk would be unattainable.

[14] In the absence of sufficient repeat sales, the Chambers, Spaenjers, and Steiner study adopts a different approach to estimating capital gains. The dataset does however allow them to comprehensively account for the impact of quality changes on rents and for costs when estimating income returns.

national price and rent index series as well as smoothing and splicing issues inherent in using historic price indices.[15] Our conclusion is that the evidence from the small number of single-country studies is nonetheless sufficient to suggest that the *total* returns to housing are lower than previously claimed.

Table 3. Comparison of Annualized Percentage Real Returns to Housing

	France 1809-1943		Holland 1900-79		UK 1901-83	
	EKLT	JKKST	EKLT	JKKST	CSS	JKKST
Capital gains	−0.6	−0.7	−0.6	0.8	−0.7	0.7
Total returns	2.8	4.2	4.8	7.1	2.3	4.7

Sources: All figures are geometric means. CSS: Chambers, Spaenjers, and Steiner (2021); EKLT: Eichholtz, Korevaar, Lindenthal, and Tallec (2021); JKKST: Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019). Note that the capital gains estimate for JKKST comes from Knoll, Schularick, and Steger (2017).

The other important conclusion upon which all four empirical studies agree is that income yield accounts for the bulk of total return. Net rental yield most likely represents the ceiling on expected total returns – and not the floor to be augmented by long-term real growth and repricing.

Our final word on real estate is an appeal for more long-run empirical studies of total returns to housing using the best available data and avoiding the measurement problems highlighted above. It is also worth emphasizing that little has been done on the long-run performance of commercial real estate, farmland, and infrastructure.[16] This is particularly important given their greater relevance to institutional investor portfolios than housing.

3.5 Commodities: Are commodities rewarded in the long run?

Many empirical studies suggest that commodities are a poor investment over the very long-run. For example, GFD reports an annual excess return over cash for an equal-weighted portfolio of diversified spot commodities of –0.4% over 1791-2020.[17] This

[15] EKLT (2021) p.3626-8 note the upward biases to JKKST capital gains estimates due to smoothing and splicing problems for both France and Holland.

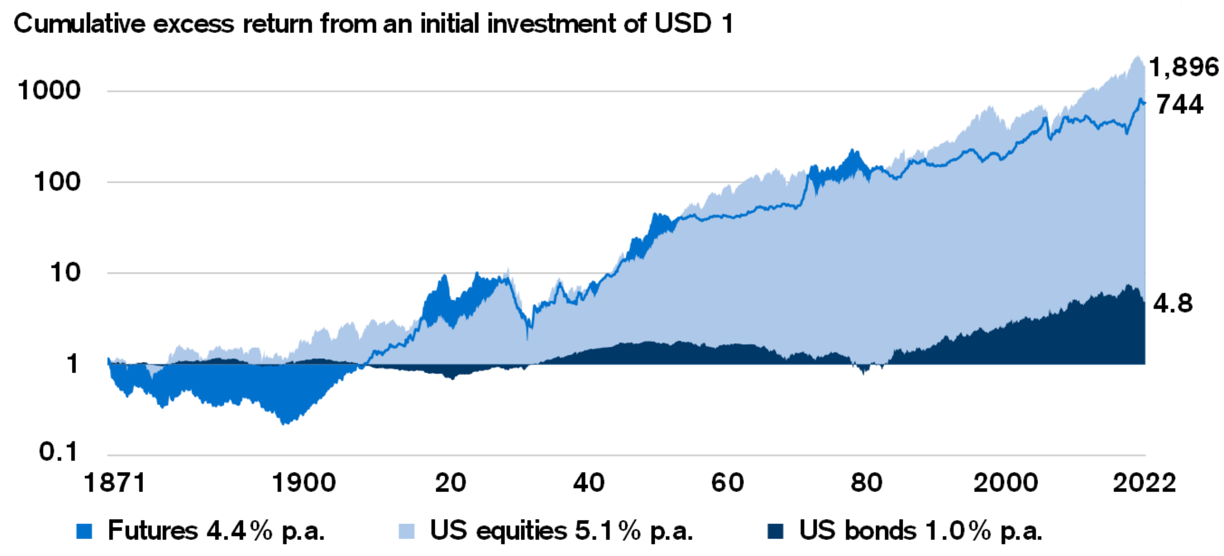
[16] Chambers Spaenjers, and Steiner (2021) report farmland total returns in England over 1901-83.

[17] For comparison, Jacks (2019) implies an annual excess return over cash of –0.2% and slightly lower real return for an equal-weighted composite of spot commodity returns between 1851 and 2015. the Economist's index of industrial spot commodities has a flat cumulative real return between 1860 and 2020. Using an even longer GFD dataset, Zaremba, Bianchi, and Mikutowski (2021) document –0.1%

can be rationalized by the fact that commodities are not income-generating assets like stocks and bonds, or by theories pointing to the commodity-price-hedging and general-inflation-hedging services, or by the Prebisch-Singer hypothesis predicting a negative relative price trend for commodities versus manufactured goods (Harvey, Kellard, Madsen and Wohar (2010)).

However, when we move away from *spot* returns, empirical evidence based on a well-diversified portfolio of commodity *futures* suggest that there is a positive commodity premium (long-term excess returns over cash). Early studies found positive rewards for equal-weighted composites of commodity futures; see Bodie and Rosansky (1980) using 1950-76 data, Erb and Harvey (2006) over 1970-2004, and Gorton and Rouwenhorst (2006) over 1959-2004. Erb and Harvey contrasted such *portfolio* evidence with the much more modest returns from *individual* commodities, explaining the gap by the so-called diversification or rebalancing return (see Booth and Fama (1992)), dubbing this phenomenon as "turning water into wine."

Figure 1. Commodity Futures, US Stocks, and US Bonds, 1871-2022



Source: Dimson, Marsh, and Staunton (2023: Chapter 8) using the equally weighted commodity futures index created by Bhardwaj, Janardanan and Rouwenhorst (2019) and updated by SummerHaven Investment Management, linking into the AQR equally-weighted commodity futures index after November 2021. The US inflation rate and equity series are from the DMS Database 2023.

annual real return over multiple centuries. Erb and Harvey (2013) find little change in the real price of gold over two millennia.

More recent studies have extended the history of commodity futures returns to the 19th century: Levine, Ooi, Richardson, and Sasseville (LORS) (2018) back to 1877, Bhardwaj, Janardanan, and Rouwenhorst (2019) back to 1871, and Geczy and Samonov (2019) back to 1877. All three studies confirmed a positive premium over cash for an equal-weighted portfolio of commodity futures from the 1870s up to the present. Whilst the datasets vary and always involve narrower universes in the early decades, the main takeaway remains the same whether one starts in the 1870s or 1900 or 1950 or 1970. Indeed, **Figure 1** shows that commodity futures almost matched the long-run returns of US equities over a 152-year window (and in fact exceeded the investment return from equities since 1900).

To better understand the nature of the commodity premium (focusing on the long-run compound return over cash), we examine its three possible sources.

First, the higher long-run returns for commodity futures over spot commodities suggest a positive long-run roll return (carry) in commodity futures. However, the only long-run futures study that splits commodity futures returns into spot and roll returns, LORS (2018), reaches mixed conclusions on their relative long-run contribution. Unlike spot commodity investments, futures investments naturally delay the cash outlay, so futures effectively earn excess-of-cash returns. The relative roles of spot and roll returns in commodity futures investing depend mechanically on whether the cash return is subtracted from the roll or spot return in performance attribution.[18]

Second, it is possible that there is a selection bias in futures contracts that are created or that survive. We find little evidence that commodities for which futures contracts have been created have higher average spot returns than those without futures. However, Bhardwaj, Janardanan and Rouwenhorst (2019) find in their broad universe of 230 contracts clear evidence of survivorship bias, in that commodity futures with shorter contract histories have lower average returns.

Last, the diversification return, whilst small for most asset classes, is a major feature within commodity portfolios, thanks to the combination of high volatilities and low correlations among constituents. The greater is the volatility drag or variance drain,

[18] The 1877-2020 compound excess return over cash of a commodity futures portfolio is 3.0% (p.a.) with monthly rebalancing (used in LORS 2018) and 4.3% with annual rebalancing (used below for consistency with GFD data). Slower annual rebalancing benefited from commodities' trending tendencies "by allowing the momentum to play out."

the more it hurts compounding and thus the geometric mean return (for a given arithmetic mean). Reducing portfolio volatility by diversification mitigates this variance drain and thus boosts the portfolio geometric mean compared to the typical single-commodity geometric mean.

Table 4 illustrates the beneficial impact of the diversification return in commodities utilising data on an equal-weighted futures portfolio.[19] This diversified portfolio generates over 1877-2020 a geometric mean excess return over cash of 4.3%, compared to a mean of 0.4% across single-commodity portfolios. This additional return reflects the volatility reduction from an average of 37% across individual commodities to 22% for the diversified portfolio. Similarly, the equal-weighted spot commodity portfolio (GFD data) has a 1.1% geometric mean excess return over cash, compared to -1.5% single-commodity mean experience, reflecting the volatility reduction from 32% to 15%. The AQR spot experience is shown to disentangle different effects.

Summarizing, the historical evidence suggests that whilst single commodities may not be expected to outperform cash in compound returns, a diversified commodity futures portfolio can offer an expected premium of 3-4% per annum. Furthermore, it has achieved this over long periods – whether the past 150, 100 or 50 years. Overall, we conclude that this diversification return is the most important and empirically robust contributor to the positive long-term commodity premium.

Table 4. The commodities diversification return: excess return over cash 1877-2020

% p.a.	AQR Futures		AQR Spot		GFD Spot	
	EW	Commodity	EW portfolio	Commodity	EW portfolio	Commodity
GM	4.3	0.4	0.3	−1.2	1.1	−1.5
AM	6.5	6.0	2.3	3.4	2.1	2.7
Vol	22.2	37.0	20.9	32.3	14.5	32.0
SR	0.29	0.16	0.11	0.10	0.14	0.08

Sources: AQR refers to the dataset used in LORS (2018) extended by AQR to 2020 using annual rebalancing. GFD refers to Global Financial Data. EW portfolio is the equal weighted portfolio return across commodities. "Commodity" is the mean statistic across commodities when each commodity is held individually. GM, AM, Vol, and SR refer to geometric mean, arithmetic mean, standard deviation, and Sharpe ratio, respectively. The reported statistics for all datasets include the whole available universe of commodities at each point in time (rising from single digits to 30 (AQR) or 46 (GFD), which explains different AMs for EW portfolios and single commodities).

[19] Equal weighting is an especially simple active strategy, as each futures contract is regularly rolled into a new contract in same size irrespective of performance. More realistic weighting schemes that incorporate asset volatility or liquidity can achieve comparable volatility reduction if their rebalancing rules ensure broad diversification.

4. CONCLUSION

The differing approaches to research on long-run asset returns summarized above deserve both credit and encouragement.

We salute empirical researchers who – using increasingly electronic "archives" – are creating superior data series of raw historical data, and who are giving careful consideration to interpreting them. Beyond pioneers like Alfred Cowles, Lawrence Fisher, James Lorie and Roger Ibbotson, we offer special and more recent hat-tips to Bryan Taylor, the creator of Global Financial Data, to John Turner and his colleagues at Queens University Belfast, and to Edward McQuarrie at the Leavey School of Business, Santa Clara University.

There is also a cohort of applied researchers who address interesting questions using very long data histories created by others. One example is the Anarkulova, Cederburg, and O'Doherty (2022) team, who deploy raw data from Global Financial Data in imaginative ways, showing that the frequency of negative long-horizon returns over almost 200 years is an order of magnitude higher when studying a global equity dataset compared to US-only equity returns, or Arnott and Bernstein (2002) who studied the components and determinants of the US equity premium.

Some researchers are combining both aspects. For example, Guido Baltussen and his colleagues have extended the CRSP dataset of US stocks back to 1866–1926 to study the robustness of cross-sectional return predictability (factor premia), and elsewhere they analyze asset class performance in different inflation environments (Baltussen, Swinkels, and van Vliet (2023)).

It is increasingly important to evaluate best practices for research on long-term financial history. The methodological contributions from Jan Annaert and his co-authors are of particular significance. We, too, emphasize the importance of good documentation of data sources and choices/assumptions, as well as more focus on possible biases.

The most obvious empirical research on US and UK equity and bond markets before 1900 has already been made. It is desirable to refine these return series by improved estimates of delisting bias, default rates, broad payout measures, etc.

Another natural path is to extend empirical research beyond these core markets to other countries and to other asset classes. Any global comparisons will require data on foreign exchange rates (if unhedged), relevant short rates (if hedged), or purchasing power parity estimates. Some analyses could be extended before 1800 following the pioneering steps of Le Bris, Goetzmann, and Pouget (2019), and others. Further microstudies on real estate may be important and feasible, given the even larger role of this asset class in investor wealth before 1900. We can hope for more histories of house prices and rental income which engage with the data challenges including the need for quality adjustments as houses age.

Further analysis of forward-looking expected returns (time series predictability), or volatility and other higher moments, can supplement analysis of historical average returns.

Finally, very long historical datasets face more severe econometric challenges. Their potential usefulness for long-horizon predictability studies is balanced by statistical problems highlighted by Boudoukh, Israel, and Richardson (2019, 2022). Similar research may help deal effectively with structural changes and missing data. Farmer, Nakamura, and Steinsson (2023) stress how pernicious gradual structural changes are since we learn about them very slowly.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review. Antti Ilmanen discloses that he is employed by AQR Capital Management and affirms that he co-authored this paper in a personal capacity.

ACKNOWLEDGMENTS

We thank Andrew Ang, Jan Annaert, Gareth Campbell, John Campbell, Mark Carhart, Rik Frehen, William Goetzmann, Amit Goyal, Campbell Harvey, Roger Ibbotson, Matti Ilmanen, Sara Jonsson, Peter Koudjis, Colin Lizieri, Paul Marsh, Edward McQuarrie, Mikko Niskanen, Peter Nyberg, Elias Ohneberg, Jarkko Peltomäki, Zhenkai Ran, Raghu Rau, Geert Rouwenhorst, Stephen Schaefer, Moritz Schularick, Jeremy Siegel, Laurence Siegel, Christophe Spaenjers, Mike Staunton, Laurens Swinkels, Bryan Taylor, Annalisa Tonetto, John Turner, Kevin van Mencxel, Stijn van Nieuwerburgh, Josef Zechner, and the editor, Matthew Richardson, for data provision and helpful conversations on this topic. We also thank seminar participants at London School of Economics, Queen’s University Belfast, Stockholm University, and the Quantitative Endowment Investing Symposium at Potsdam, Germany.

LITERATURE CITED

1. Acheson, Graeme G., Charles R. Hickson, John D. Turner, and Qing Ye (2009). Rule Britannia! British stock market returns, 1825-1870, Journal of Economic History 69, 1107–1137.
2. Amaral, Francisco, Martin Dohmen, Sebastian Kohl, and Moritz Schularick (2023). Superstar returns. Journal of Finance **R&R**.
3. Anarkulova, Aizhan, Scott Cederburg, and Michael S. O'Doherty (2022). Stocks for the long run? Evidence from a broad sample of developed markets. Journal of Financial Economics, 143(1): 409-433.
4. Annaert, Jan, Frans Buelens, and Angelo Riva (2016). Financial history databases: old data, old issues, new insights, in Chambers, David, and Elroy Dimson [Eds.] (2016) Financial Market History, 44-58.
5. Arnott, Robert, and Peter Bernstein (2002). What risk premium is normal? Financial Analysts Journal, 58(2): 64-85.
6. Baltussen, Guido, Laurens Swinkels, and Pim van Vliet (2023). Investing in deflation, inflation, and stagflation regimes, Financial Analysts Journal, 79(3): 5-32.
7. Baltussen, Guido, Bart van Vliet, and Pim van Vliet (2023). The cross-section of stock returns before CRSP. Working paper (ssrn.com/id=396974).
8. Bernstein, William (2022). Review of Antti Ilmanen: Investing Amid Low Expected Returns. Advisor Perspectives.
9. Bhardwaj, Geetesh, Rajkumar Janardanan, and K. Geert Rouwenhorst (2019). The commodity futures risk premium: 1871–2018. Working paper (ssrn.com/id=3452255).
10. Bodie, Zvi, and Rosansky, Victor (1980). Risk and return in commodity futures. Financial Analysts Journal, 36(3): 27-39.
11. Booth, David G., and Eugene F. Fama (1992). Diversification returns and asset contributions, Financial Analysts Journal 48(3): 26-32.
12. Boudoukh, Jacob, Ronen Israel, and Matthew Richardson (2019). Long-horizon predictability: A cautionary tale, Financial Analysts Journal 75(1): 17-30.
13. Boudoukh, Jacob, Ronen Israel, and Matthew Richardson (2022). Biases in long-horizon predictive regressions, Journal of Financial Economics 145(3): 937-969.
14. Brennan, Michael J., and Eduardo S. Schwartz (1985). On the geometric mean index: A note, Journal of Financial and Quantitative Analysis, 20(1): 119-122.
15. Brown, Stephen J., William N. Goetzmann, and Stephen A. Ross (1995). Survival. Journal of Finance 50(3): 853-873.
16. Campbell, Gareth, Richard Grossman, and John D. Turner (2021). Before the cult of equity: the British stock market, 1829–1929. European Review of Economic History, 25(4): 645-679.
17. Campbell, Gareth, William Quinn, John D. Turner, and Quing Ye (2018). What moved share prices in the nineteenth-century London stock market? Economic History Review, 71(1):157-189.
18. Campbell, John Y. (2018). Financial Decisions: A Course in Asset Pricing. Princeton University Press.
19. Chambers, David, and Elroy Dimson [Eds] (2016). Financial Market History: Reflections on the Past for Investors Today. CFA Institute Research Foundation.
20. Chambers, David, Christophe Spaenjers, and Eva Steiner (2021). The rate of return on real estate: Long-run micro-level evidence, Review of Financial Studies, 34(8), 3572-3607.
21. Chancellor, Edward (2022). The Price of Time. The Real Story of Interest. Atlantic Monthly Press.
22. Cochrane, John H. (2011). Presidential address: Discount rates. Journal of Finance 66:1047–108.
23. Cowles, III, Alfred, and Associates (1938). Common Stock Indices, 1871-1937, Cowles Commission Monograph no. 3. Bloomington, Ind.: Principia Press.
24. Coyle, Christopher, and John D. Turner (2013). Law, politics, and financial development: the great reversal of the U.K. corporate debt market, Journal of Economic History 73: 810–846.
25. Dimson, Elroy, Paul Marsh, and Mike Staunton (2002). Triumph of the Optimists: 101 Years of Global Investment Returns, Princeton University Press.
26. Dimson, Elroy, Paul Marsh, and Mike Staunton (2018). Private wealth investments. Chapter 4 of Global Investment Returns Yearbook 2018: 64-82. Zurich: Credit Suisse Research Institute.
27. Dimson, Elroy, Paul Marsh, and Mike Staunton (2021). American exceptionalism. Journal of Portfolio Management 47(7): 14–26.

28. Dimson, Elroy, Paul Marsh, and Mike Staunton (2023). Global Investment Returns Yearbook 2023. Zurich: Credit Suisse Research Institute.
29. Eichholtz, Piet, Matthijs Korevaar, Thies Lindenthal, and Ronan Tallec (2021). The total return and risk to residential real estate, Review of Financial Studies 34(8): 3608-3646.
30. Erb, Claude B., and Campbell R. Harvey (2006). The tactical and strategic value of commodity futures, Financial Analysts Journal 62(2): 69-97.
31. Erb, Claude B., and Campbell R. Harvey (2013). The golden dilemma, Financial Analysts Journal 69(4): 10-42.
32. Eytan, Hanan, and Giora Harpaz (1986). The pricing of futures and options contracts on the Value Line Index, Journal of Finance, 43(3): 843-855.
33. Farmer, Leland E., Emi Nakamura, and Jon Steinsson (2023), Learning About the Long Run, forthcoming in the Journal of Political Economy.
34. Fisher, Lawrence, and James H. Lorie (1964). Rates of return on investments in common stocks. Journal of Business. 37(1): 1-21.
35. Fisher, Lawrence, and James H. Lorie (1968). Rates of return on investments in common stock: the year-by-year record, 1926-65. Journal of Business, 41(3): 291-316.
36. Foreman-Peck, James, and Leslie Hannah (2012). Extreme divorce: the managerial revolution in UK companies before 1914. Economic History Review, 65: 1217-1238.
37. Gayer, Arthur, Anna Jacobson, and Isaiah Finkelstein (1940). British share prices, 1811-1850. Review of Economics and Statistics 22(2): 78-93.
38. Geczy, Christopher C., and Mikhail Samonov (2019). Two centuries of commodity futures premia: momentum, value and basis. Working paper (ssrn.com/id=3336406)
39. Giesecke, Kay, Francis A. Longstaff, Stephen Schaefer, and Ilya Strebulaev (2011). Corporate bond default risk: a 150-year perspective, Journal of Financial Economics 102, 233–250.
40. Goetzmann, William N., Roger G. Ibbotson, and Liang Peng (2001). A new historical database for the NYSE 1815 to 1925: performance and predictability. Journal of Financial Markets, 4(1): 1-32.
41. Goetzmann, William N., Christophe Spaenjers, and Stijn Van Nieuwerburgh (2021). Real and private-value assets. Review of Financial Studies 34(8): 3497–3526.
42. Golez, Benjamin, and Peter Koudijs (2018). Four centuries of return predictability. Journal of Financial Economics 127: 248–263.
43. Gorton, Gary B., and K. Geert Rouwenhorst (2006). Facts and fantasies about commodity futures, Financial Analysts Journal 62: 86-93.
44. Graham, Benjamin, and Jason Zweig (2024). The Intelligent Investor. Revised Edition, forthcoming.
45. Gregory, Alan (2023). The expected cost of equity in the UK revisited. Working paper (ssrn.com/id=4516096).
46. Grossman, Richard (2002). New indices of British equity prices, 1870–1913, Journal of Economic History 62(1): 121–146.
47. Gyourko, Joseph, Christopher Mayer, and Todd Sinai. (2013). Superstar cities. American Economic Journal: Economic Policy, 5 (4): 167-99.
48. Harvey, David, Neil Kellard, Jakob Madsen, and Mark Wohar (2010). The Prebisch–Singer hypothesis: four centuries of evidence. Review of Economics and Statistics. 92(2): 367–377.
49. Hautcoeur, Pierre-Cyrille, and Angelo Riva. (2018). The Data for Financial History (DFIH) Database. Paris: Paris School of Economics.
50. Hickman, W. Braddock (1960). Statistical measures of corporate bond financing since 1900, National Bureau of Economic Research, Studies in Corporate Bond Financing. Princeton University Press, Princeton.
51. Homer, Sidney, and Richard Sylla (2005), A History of Interest Rates, Rutgers University Press.
52. Ibbotson, Roger G., and Rex A. Sinquefield (1976). Stocks, bonds, bills and inflation: year-by-year historical returns, Journal of Business 49: 11-47.
53. Ilmanen, Antti (2022). Investing Amid Low Expected Returns, Wiley.
54. Jacks, David S. (2019). From boom to bust: a typology of real commodity prices in the long run. Cliometrica 13(2): 202-220.
55. Jordà Òscar;, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M Taylor (2019). The rate of return on everything, 1870–2015, Quarterly Journal of Economics 134(3): 1225–1298.
56. Jorion, Philippe, and William N. Goetzmann (1999). Global stock markets in the twentieth century. Journal of Finance, 54(3):.953-980.

57. Kizer, Jared, Sean Grover, and Corey Hendershot (2019). Re-examining the credit premium, Working paper (ssrn.com/id=3147005).
58. Klovland, Jan Tore (1994). Pitfalls in the Estimation of the Yield on British Consols, 1850-1914. Journal of Economic History 54(1):164-187.
59. Knoll, Katharina (2017). Our Home in Days Gone By: Housing Markets in Advanced Economies in Historical Perspective. PhD thesis, Free University of Berlin.
60. Knoll, Katharina, Moritz Schularick, and Thomas Steger (2017). No price like home: Global house prices, 1870-2012, American Economic Review 107(2): 331-53.
61. Korevaar, Matthijs (2023). Reaching for yield: How investors amplify housing booms and busts. Working paper (ssrn.com/id=3794782).
62. Kuvshinov, Dmitry, and Kaspar Zimmermann (2022). The Big Bang: Stock Market Capitalization in the Long Run. Journal of Financial Economics 145(2): 527-552.
63. Le Bris, David, William N. Goetzmann, and Sébastien Pouget (2019). The present value relation over six centuries: The case of the Bazacle company, Journal of Financial Economics 132(1): 248-265.
64. Le Bris, David, and Pierre-Cyrille Hautcœur (2010). A challenge to triumphant optimists? A blue chips index for the Paris stock exchange, 1854–2007, Financial History Review 17(2): 141–183.
65. Lettau, Martin, Sydney C. Ludvigson, and Jessica A. Wachter (2008). The declining equity premium: what role does macroeconomic risk play? Review of Financial Studies 21(4): 1653-1687.
66. Levine, Ari, Yao Hua Ooi, Matthew Richardson, and Caroline Sasseville (2018). Commodities for the long run, Financial Analysts Journal 74(2): 55-68.
67. Lyons, Ronan, Allison Shertzer, Rowena Gray and David Agorastos (2023). The Price of Housing in the United States, 1890-2006. Working Paper.et al
68. Macaulay, Frederick R. (1938). Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields and Stock Prices in the United States Since 1856. Journal of the Royal Statistical Society, 101(3): 620–623.
69. Maddison – Groningen Database (2020). Groningen Growth and Development Centre. (www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020).
70. Marks, Peter, and Alan Stuart (1971). An arithmetic version of the Financial Times Industrial Ordinary Share Index. Journal of the Institute of Actuaries, 97(2-3): 297–324.
71. McQuarrie, Edward F. (2023). Stocks for the long run? Sometimes Yes. Sometimes No. Financial Analysts Journal **R&R**.
72. McQuarrie, Edward F. (2020). When do corporate bond investors earn a premium for bearing risk? a test spanning the great depression of the 1930s. Working paper (ssrn.com/id=3740190).
73. McQuarrie, Edward F. (2021). The US bond market before 1926: Investor total return from 1793, comparing federal, municipal, and corporate bonds Part I: 1793 to 1857 and Part II: 1857 to 1926. Working papers (ssrn.com/id=3260733 and ssrn.com/id=3269683).
74. Meyer, Josefin, Carmen M. Reinhart, and Christoph Trebesch (2022). Sovereign bonds since Waterloo. Quarterly Journal of Economics 137(3): 1615-1680.
75. Pástor, Ľuboš, and Robert F. Stambaugh (2001). The equity premium and structural breaks, Journal of Finance 56(4): 1207-1239.
76. Piketty, Thomas (2013). Capital in the Twenty First Century, Harvard University Press.
77. Schmelzing, Paul (2020), Eight centuries of global real interest rates, R-G, and the 'suprasecular' decline, 1311–2018, Bank of England Working Paper No. 845 (ssrn.com/id=3520654).
78. Schwert, G. William (1990). Indexes of US stock prices from 1802 to 1987, Journal of Business 63(3): 399-426.
79. Sharpe, William F. (2010). Adaptive asset allocation policies. Financial Analysts Journal 66:45–59.
80. Shiller, Robert, (2005). *Irrational Exuberance*. Princeton University Press, 2nd edition.
81. Shumway, Tyler (1997). The Delisting Bias in CRSP Data. Journal of Finance, 52(1): 327–340.
82. Siegel, Jeremy J. (1992). The Equity Premium: Stock and Bond Returns Since 1802. Financial Analysts Journal, 48(1): 28-46.
83. Siegel, Jeremy J. (1994, 2014, 2022). Stocks for the Long Run, 1st, 5th and 6th editions, McGraw Hill.
84. Siegel, Laurence, and Paul McCaffrey [Eds] (2023). Revisiting the Equity Risk Premium. CFA Institute Research Foundation.

85. Smith, W. B. and Cole, A. H., (1935). Fluctuations in American Business 1790-1860. Russell & Russell.
86. Taylor, Bryan (2023). GFD Guide to Global Stock Markets, 2023 Edition. Global Financial Data. Working paper (ssrn.com/id=4514486).
87. Thomas, Sam (2002). The Saga of the first stock index futures contract: benchmarks, models, and learning. Journal of Money, Credit and Banking, 34(3): 767–808.
88. Thomas, Ryland, and Nicholas Dimsdale (2018). A millennium of macroeconomic data for the UK. Bank of England.
89. Turnbull, Craig (2017). A history of British actuarial thought. Palgrave Macmillan.
90. Van Mencxel, Kevin, Jan Annaert, and Marc Deloof (2022). Bonds for the long run? The rate of return on corporate bonds in Belgium, 1838–1939. Working paper (ssrn.com/id=4120896).
91. Wicksell, Knut (1898) [trans. Kahn, R.F., 1936]. Interest and prices. United States: Creative Media Partners, LLC.
92. Woodford, Michael (2003). Interest and Prices, Princeton University Press.
93. Zaremba, Adam, Robert J. Bianchi, and Mateusz Mikutowski (2021). Long-run reversal in commodity returns: Insights from seven centuries of evidence, Journal of Banking & Finance 133(7).
94. Zweig, Jason (2009). Does stock-market data really go back 200 years? Wall Street Journal (July 11th).